Journal of Physics: Conference Series 2017 vol.913 N1

Intelligent search in Big Data

Birialtsev E., Bukharaev N., Gusenkov A.
Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© Published under licence by IOP Publishing Ltd. An approach to data integration, aimed on the ontology-based intelligent search in Big Data, is considered in the case when information objects are represented in the form of relational databases (RDB), structurally marked by their schemes. The source of information for constructing an ontology and, later on, the organization of the search are texts in natural language, treated as semi-structured data. For the RDBs, these are comments on the names of tables and their attributes. Formal definition of RDBs integration model in terms of ontologies is given. Within framework of the model universal RDB representation ontology, oil production subject domain ontology and linguistic thesaurus of subject domain language are built. Technique of automatic SQL queries generation for subject domain specialists is proposed. On the base of it, information system for TATNEFT oil-producing company RDBs was implemented. Exploitation of the system showed good relevance with majority of queries.

http://dx.doi.org/10.1088/1742-6596/913/1/012010

References

[1] Hopkins B and Evelson B 2012 Expand your digital horizon with Big Data Avaliable from: http://www.asterdata.com/newsletter-images/30-04-2012/resources/forrester-expand-your-digital-horiz.pdf - ref-separator -
[2] NOSQL Databases Avaliable from: http://nosql-database.org/ - ref-separator -
[3] Apache Hadoop, MapReduce Avaliable from: https://hadoop.apache.org/docs/r1.2.1/map-red-tutorial.html - ref-separator -
[4] Apache Hadoop Avaliable from: http://hadoop.apache.org/ - ref-separator -
[5] SAP HANA Avaliable from: https://www.sap.com/products/hana.html - ref-separator -
[6] Kogalovskij M 2010 Methods of data integration in information systems (Moscow: Russian Academy of Science Institute of Market Problems) Avaliable from: http://www.ipr-ras.ru/articles/kogalov10-05.pdf
[7] W3C, OWL Web Ontology Language Avaliable from: https://www.w3.org/TR/2004/REC-owl-features-20040210/ - ref-separator -
[8] Birialtsev E, Gusenkov A and Kosinov Y 2007 Proc. of Kazan School on Computer and Cognitive Linguistics TEL-2006 (Kazan: Otechestvo Publishing House) Representation of relational database structure in the ontology formalism 32-37
[9] Birialtsev E and Gusenkov A 2007 Integration of databases based on ontologies Scholarship Notes of the Kazan University. Physics and mathematics series 149 13-25
[10] Gusenkov A, Birialtsev E and Zhibrik O 2015 Intellectual search in structured information arrays (LAP LAMBERT Academic Publishing)
[11] Garcia-Molina H, Ullman J and Widom J 2003 Database Systems: The Complete Book (Williams Publishing House)
[12] Booch G, Rumbaugh J and Jacobson I 2000 The Complete UML Training Course (Moscow: DMK Press)

[13] Birialtsev E, Gusenkov A and Galimov M 2005 Features of the lexico-semantic structure of the relational databases artifacts naming Proc. of Kazan School on Computer and Cognitive Linguistics TEL-2005 4-12

[14] Zhuchkov A, Arnautov S and Tverdohlebov N 2003 New technologies for conceptual networks created within framework of the project "New generation vaccines and diagnostic systems of the future" Electronic Libraries 6 Avaliable from: http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2003/part6/ZATGS

[15] Gavrilova T and Horoshevskij V 2001 Knowledge base of intelligent systems (St. Petersburg: Piter Publishing House)

[16] Anderson J 2003 Discrete Mathematics with Combinatorics (Moscow: Williams Publishing House)

[17] Birialtsev E, Gusenkov A and Khairullina A 2006 Presentation of the POSC Epicenter data model in the OWL ontology language Proc. of Kazan School on Computer and Cognitive Linguistics TEL-2006 38-49

[18] Birialtsev E and Gusenkov A 2007 Proc. of Knowledge-Ontology-Theory (ZONT-07) Conf. (Novosibirsk: Sobolev Mathematical Institute Press 1) The construction of subject domain ontology based on the logical database model 176-183

[19] Energistics, Epicentre v3.0 Avaliable from: http://www.energistics.org/energistics-standards-directory/epicentre-archive - ref-separator -

[20] Fellbaum C (ed) 1998 WordNet: An Electronic Lexical Database (Cambridge: MIT Press)

[21] Birialtsev E and Gusenkov A 2007 Relational database ontologies. Linguistic aspect Proc. Of Dialog'2007 Conf. 50-53

[22] Narinyani A S 2001 Centaur named Theon: Thesaurus + Ontology Proc. of the international workshop on computational linguistics and its applications DIALOG 1 184-188

[23] Birialtsev E, Gusenkov A and Mironov S 2009 One approach to implementing unregulated access to relational databases Proc. of Kazan School on Computer and Cognitive Linguistics TEL-2008 10-23

[24] Chaudhuri S and Dayal U 1997 An Overview of Data Warehousing and OLAP Technology ACM SIGMOD Record 26 65-74